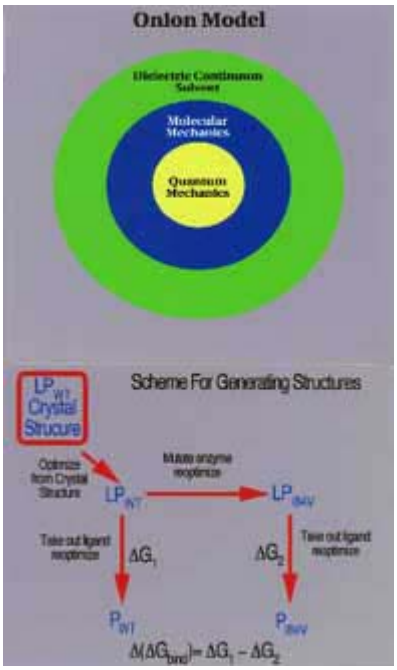
Determination of Enzyme-Substrate Binding Affinities

A Research Project of Dr. Gregory J. Tawa and Dr. Igor A. Topol

Introduction



Theoretical approaches that assess ligand-protein binding affinity prior to synthesis and testing of ligands are of obvious importance in the field of structure based drug design. However, an understanding of the principles of ligand-protein binding thermodynamics and the computation of ligand-protein binding affinities are difficult problems for which there is currently no satisfactory solution. At the most general level, our research addresses each of these issues (principles and computation) that we feel are an important contribution in the area of structure based drug design. One particular area of research is understanding drug inhibition of HIV protease, an important therapeutic target in the treatment of AIDS. HIV-1 protease is responsible for the post-translational processing of the polyprotein gene products of gag and gag-pol to yield the structural proteins and enzymes of the viral particle, and is an essential element in the infection process. Emergence of drug-resistant variants (mutants) of HIV in response to exposure to various inhibitors has severely limited the effectiveness of even the most promising drugs. Our research strives to understand the principles of resistance by computation and analysis of the change in thermodynamic variables relevant to drug-protease binding. The drug resistance problem is not only limited to AIDS. Recently, evolved drug-resistant strains of the tuberculosis bacillus have been plaguing industrial urban centers. The theoretical approaches developed at the ABCC for describing resistance in HIV-1 will also have bearing on this system as well as others for which mutation hinders the effectiveness of drug therapies.

Methodologies

In aqueous solution, a ligand (L) and a protein (P) associate to form a ligand-protein complex (LP). The binding affinity or absolute binding free energy, ΔGb, is given by

$$\Delta G_b(LP) = G_{aq}(LP) - G_{aq}(L) - G_{aq}(P), \quad (1)$$

and relative binding free energy of two closely related proteins (different by only one residue) to a ligand is given by

$$\Delta(\Delta G_b) = [G_{aq}(LP_2) - G_{aq}(LP_1)] + [G_{aq}(P_1) - G_{aq}(P_2)], \quad (2)$$

where G_{aq} is the aqueous phase free energy. In our modeling we treat the enzyme-ligand-solvent complex with an onion type model. In this model the inner most circle represents the enzyme active site region (which will also include the ligand in the case of ligand-protein complexes). This subsystem is treated in molecular detail using quantum mechanics. The next circle out is the rest of the protein, which is treated with standard molecular mechanics. The outermost region is the solvent, which is represented as a structureless polarizable dielectric continuum. Quantum mechanical calculations are performed first on the inner-most subsystem to determine the protonation state of the active site residues of all species present in Equation 2. Given the protonation states, the complete protein and ligand-protein complexes are optimized (starting with published crystal structure coordinates) using standard molecular mechanics potentials to obtain the gas phase energies (E) and the geometries. All species are then hydrated. At this stage of the calculation the protein and ligand-protein complexes are represented by a set of atomic charges (from molecular mechanics database) centered on the

Table 1. Gas phase and solvation components of the change in binding free energy due to mutation, $\Delta(\Delta Gb)$, for Ro318959, L735-524, KNI-272, and A-77003 binding to wild-type HIV-1 protease and its I84V mutant.

Ro31-8959	
$\Delta(\Delta E)^a$	1.05
$\Delta(\Delta[\Delta G_{el}])^b$	-0.06
$\Delta(\Delta G_b)^c$	0.99 1.04[d]
L735-542	
$\Delta(\Delta E)^a$	-0.66
$\Delta(\Delta[\Delta G_{el}])^b$	3.06
$\Delta(\Delta D_b)^c$	2.40 1.36[d]
A-77003	
$\Delta(\Delta E)^a$	5.58
$\Delta(\Delta[\Delta G_{el}])^b$	-3.68
$\Delta(\Delta G_b)^c$	1.90 1.30[d]
KNI-272	
$\Delta(\Delta E)^a$	0.39
$\Delta(\Delta[\Delta G_{el}])^b$	1.64
$\Delta(\Delta G_b)^c$	2.03 2.05[d]

[a] $\Delta(\Delta E) = [E(LP_2) - E(LP_1)] + [E(P_1) - E(P_2)]$

atoms of the gas phase optimized structures. This protein charge distribution exerts an electric field out into the solvent. The solvent becomes polarized (this is represented by polarization charges that develop on the van der Waals surface) and the resultant Coulomb interaction between the solvent polarization and the protein charge density defines the electrostatic hydration free energy DGel. The full aqueous phase free energy of each species is then approximated by

$$G_{aq} = E + \Delta G_{el} \qquad (3)$$

and these are used in Equation 2 to obtain the relative binding free energies. The optimization protocol is shown in the scheme below. This procedure is designed so that all calculations are confined to a limited region of configuration space. This insures that the double subtraction in the energies will have the appropriate cancellation errors.

[b] $\Delta(\Delta[\Delta G_{el}]) = [\Delta G_{el}]) = [\Delta G_{el}(LP_2) - \Delta G_{el}(LP_1)] + \{\Delta G_{el}(P_1) - \Delta G_{el}(P_2)]$

[c] Eq. (2)

[d] Experiment: Dr. Sergei Gulnik, SBP

Application to Calculating Accurate Binding Energies

We have considered the binding of inhibitors KNI-272, Ro31-8959,L735-524, and A-77003 to HIV-1 protease and its I84V mutant. The calculated and experimental relative binding free energies, D(DGb), are given in Table 1. These are partitioned into gas phase energetic contributions, D(DE), and electrostatic hydration contributions, D(D[DGel]). The calculated relative binding free energies exhibit an average deviation of 0.39 kcal/mol from experiment. In the case of Ro31-8959 and KNI-272, the calculated results are essentially identical to experiment. The largest deviations from experiment occur for A-77003 and L735-524, these are +0.6 and +1.04 kcal/mol respectively. However, these deviations are roughly the same size as the experimental error bars. In short, the theoretical results match the experimental ones rather well. Analysis of the gas phase and hydration components of the relative binding free energy across inhibitors shows that either component alone does not correlate with the experimental relative binding free energies. However, their sum does. Furthermore, the hydration component always corrects the gas phase relative binding free energy in a direction closer to the experimental values. Clearly, quantitative accuracy in the relative binding free energies can only be obtained if hydration effects are included. By analyzing the energetic and hydration components of the relative binding free energies, we discover three types of binding scenarios. In the case of Ro31-8959 the hydration component of the relative biding free energy is small compared to the gas phase component; therefore, the loss in binding affinity due to mutation is determined mainly by the change in their intersolute enthalpic interactions (here the solutes are defined as either the isolated enzymes, or the inhibitor-enzyme complexes). For A-77003 and L735-524 both the gas phase and the hydration components are large. The loss in binding affinity due to mutation is then defined by the change of both solute-solute and solute-solvent interactions. For KNI-272, the hydration component is large relative to the has phase component, therefore loss in binding affinity is primarily determined by the change in the solute-solvent interaction due to mutation.

The Future

The I84V mutant enzyme is only one of many that may occur. In the future we will use this methodology to obtain theoretical

computational tools (ab initio, molecular mechanics, dielectric solvation) it is ideally suited for a heterogeneous computing environment. Analysis of the various parts of the method reveals that the quantum mechanical portion of the calculation runs best on the vector pipeline architecture of the SGI/Cray J90. Currently, only a limited size active site region (at most 200 atoms with minimal bases sets) can be treated due to memory limitations. Although this has been sufficient for determining protonation states, a complete treatment of the active site would be needed to describe the mechanics of bond making and breaking that occur during substrate binding and cleavage. A quantum mechanical treatment using on the order of 200 active site atoms with large basis sets containing polarization and diffuse functions is necessary for an accurate characterization of these processes. This is currently beyond the scape of the current technology at the ABCC.

The outer protein polarization and hydration calculations involve an iterative solution of the Poisson equation to obtain induced protein dipoles (located in the outer protein) and solvent polarization charges (location on the protein van der Waals surface). This iteration is for a system of equations typically on the order of 100,000 or more. Using a Jacobi update scheme, each element of the solution vector can be updated independently of the others. This activity is ideally suited for parallel processing (i.e., the SGI/Cray Origin 2000). Construction of the matrix elements associated with the system of equations is inherently a scalar process and is best suited to run on the SGI Power Challenge. The key to putting all of these elements together into a powerful computational tool is in the high speed links and software that will allow the various computer platforms to communicate. At present, neither the software of the high speed links currently exist at the ABCC to perform these calculations.

A correlated ab initio quantum mechanical treatment using on the order of 200 or more active site atoms with large basis sets containing polarization and diffuse functions is necessary for an accurate characterization of ligand-binding processes. The iterative solution of the Poisson equation to determine polarization requires four days to CPU on an SGI Power Challenge. This requires a solution of 80,000 coupled equations. Analysis has shown that in order to obtain absolute binding free energies to within one kcal/mol precision it will be necessary to solve 500,000 to one million coupled equations. In order to accomplish this, a one or two order of magnitude increase in computational power will be required.